

Практический опыт переноса Netware сервера на новые диски

Оглавление

Вступление.....	2
Подготовка к копированию.....	3
Выбор программного обеспечения для клонирования.....	4
Копируем	5
Небольшое послесловие	7

Вступление

Жил был старый сервер (интеловская Madrona MB440LX) в однопроцессорном варианте, с двухканальным интегрированным RAID-контроллером AMI MegaRAID, ОЗУ – 256Мб). Корпус имеет две «корзинки» для hot-swap дисков с разъёмом SCA-2, по пять позиций на каждом «этаже». Реально было подключено четыре 9-ти гигабайтных диска, по два на «этаж», из которых был собран RAID-5 массив с итоговой ёмкостью ~27 Гигабайт. Работала (и работает) на сервере ОС NW4.11+sp9. С какого-то времени стала ощущаться заметная нехватка дискового пространства, которая до некоторого момента не напрягала, поскольку в «корзинах» оставались свободными ещё шесть позиций. Мыслилось поступить так – докупается шесть дисков нужного объёма (36 или 73 Гигабайта), из них средствами MegaRAID собирается ещё один логический диск, а дальше новая ёмкость распределяется по потребности по нужным томам.

Первое нехорошее предчувствие, что таким способом поступить не удастся, наступило тогда, когда выяснилось¹, что старый MegaRAID не понимает современных SCSI-дисков! При установке диска стандарта Ultra160 (SCSI-3) и новее (Ultra320) контроллер просто замирал на несколько секунд при сканировании определённой позиции и... не обнаруживал диск. Вот тебе и хваленая совместимость SCSI-оборудования! Стандартные в таких случаях мероприятия и советы по обновлению прошивки (firmware) RAID-контроллера не годились – прошивка и так была самая последняя из доступных на сайте Интела. С отдельными MegaRAID-контроллерами (в частности, их активно использовал HP в своих серверах) таких проблем не существовало – обновление прошивок для поддержки современных дисков было доступно очень давно. Интел же объединил в одну общую прошивку и BIOS для «материнской платы» и firmware для интегрированного RAID-контроллера, прекратив развитие для данной модели сервера в мае 1998 года. Переписка со службой техподдержки Интел подтвердила факт отсутствия возможности работать с новыми дисками и, главное – отсутствие желания у разработчиков исправить ситуацию. Таким образом необходимость переезда на более современный дисковый массив увеличенного объёма стала неотвратимой (перенос сервера полностью на другое «железо» не рассматривался – узким местом была только недостаточная ёмкость дисков, по всем остальным параметрам сервер устраивал нашу бухгалтерию полностью).

Неприятным свойством интегрированного MegaRAID-а было то, что конфигурация массива хранилась только во флеш-памяти устройства, а не на дисках, как сейчас практикуется во всех современных RAID-контроллерах хорошего класса (это позволяет, например, без проблем перенести собранные в массив диски на другой контроллер/сервер в случае выхода из строя оборудования на исходной машине). Эта особенность усугублялась ещё отсутствием возможности сохранить конфигурацию контроллера на внешнем носителе, типа дискеты (как это умеют делать, к примеру, устройства фирмы Mylex). Осознание этого факта приводило к необходимости крайней осторожности при манипулировании дисками, т.к. неизвестно было, к примеру, как себя поведёт MegaRAID, если перед загрузкой отключить от него SCSI-кабели, идущие к «корзинам» с дисками. Не потеряет ли он после этого массив полностью и окончательно? (И до сих пор не знаю, что в этом случае будет, на «живой» системе экспериментировать не хотелось).

¹ На самом деле это выяснилось чуть раньше, на другом, но точно таком же «железе», когда «вылетевший» старый SCSI диск безуспешно пытались заменить на более современный – контроллер не «видел» его ни в какую. В результате «длинной рокировки» нужную модель диска подобрать удалось, но возникло стойкое ощущение грядущих неприятностей J.

Поняв, что со старым RAID-контроллером не удастся добиться желаемого результата (увеличения ёмкости простой установкой новых накопителей), да и надёжность отработавших по шесть лет дисков уже вызывала некоторые опасения, было решено перейти на другой тип контроллера. После недолгих поисков был обнаружен одноканальный Adaptec SCSI RAID 2100S, имевший поддержку в виде драйверов для Netware 4.11/4.2.

Подготовка к копированию

Первым этапом переезда на новые диски был перенос всех старых дисков в имеющемся массиве на один «этаж» (при этом вторая «корзина» остаётся полностью пустой). Эта функция переноса («путешествие», *roaming*) дисков поддерживается MegaRAID-контроллером. Небольшая тонкость: в связи с описанной выше особенностью хранения конфигурации только во флеш-памяти, и во избежание преждевременной «кончины» рабочего массива, диски переносились по одному за раз. Т.е. сначала в выключенном состоянии из нижней «корзины» в верхнюю был переставлен один диск, после чего была произведена загрузка сервера. BIOS RAID-а остановился на тесте, сообщив, что обнаружил перемещение диска в другую позицию и требуя подтверждения² на продолжение загрузки. Затем после загрузки сервер был ещё раз остановлен, выключен, и на верхний «этаж» перебрался последний оставшийся диск (с последующей пробной загрузкой машины).

После переноса всех четырёх дисков на один канал RAID-контроллера сервер был в очередной раз выключен, в него была вставлена плата Adaptec SCSI RAID 2100S, а к ней подключён шлейф от освободившейся нижней «корзины». В нижний «этаж» было вставлено пять заранее собранных в массив RAID-5 SCSI дисков ёмкостью по 36 Гигабайт. Здесь сказалось преимущество более современных RAID-контроллеров, хранящих информацию о массиве на самих дисках, что дало возможность провести заранее весьма длительную процедуру построения массива (скорость «сборки» была около процента в минуту) на другой машине.

При попытке последующей загрузки (со старого массива) выяснился интересный момент: RAID-контроллер Adaptec 2100S по приоритету оказался стоящим **раньше** MegaRAID-а, поэтому вместо загрузки на экране возникла ругань на тему *non bootable device*³. Что ж, перезагружаемся, ждём появления на экране приглашения *setup BIOS-a 2100S* с предложением нажать *Ctrl-A*, нажимаем и... зависаем. Хм, странно, в чём дело? «Набортная» индикация 2100S (весьма развесистая) показывает некоторую активность, но индикаторы новых дисков с нижнего «этажа» практически безмолвствуют. Ждём минуту, две, три, четыре... оп-па – наконец-то появилось окно настроек RAID 2100S. А ведь на другой машине (правда, гораздо более современной) никаких задержек при входе в *setup BIOS-a 2100S* не наблюдалось. Читаю сопроводительные *readme* файлы к 2100S и натываюсь на упоминание, что у этого контроллера с некоторыми моделями интеловских серверных плат возможны долгие задержки при входе в *setup*. Смотрю список, обнаруживаю там нашу MB440LX. Ага, понятно, как говорится, «описанные ошибки становятся особенностью программного обеспечения», т.е., волноваться незачем, задержка входа в *setup* документирована и я её ощутил по полной (как мне тогда показалось) программе **J**.

Наконец, делаю то, ради чего так долго пришлось ждать окно с настройками RAID 2100S: нахожу соответствующий пункт меню и запрещаю новому RAID-у быть загрузабельным

² Совершенно ненужный и странный сервис. Правильнее было бы, как делают большинство современных контроллеров, просто оповестить пользователя об обнаруженном перемещении и грузиться дальше.

³ Что естественно, т.к. сервер пытался загрузиться с пустого свеже созданного массива.

(bootable) устройством. Потом, по окончании клонирования, надо будет не забыть вернуть эту птичку на место, чтобы грузиться уже с нового массива.

Итак, имеем: два RAID-5 массива – старый и новый (пустой) – физически подключены к серверу, теперь дело за программным обеспечением, позволяющим перенести данные с одного массива на другой.

Выбор программного обеспечения для клонирования

Выбор программного обеспечения для копирования (клонирования) одного массива на другой для ОС Netware не очень велик, на сегодня это фактически две программы: Norton GHOST for Netware v2 (ничего не знает про NSS) и Portlock Storage Manager. У меня был успешный опыт работы с первой из них (NG4NW), и вначале я собирался воспользоваться именно этой программой. Её особенность – она работает только в DOS-е (правда требует много ОЗУ, необходимого ей для монтирования томов). Но тут выяснился некоторый интересный момент, напроць перечеркнувший возможность использовать для моей конфигурации NG4NW: многие RAID-контроллеры для DOS-а отдают не всю реально доступную ёмкость массива, а некоторую гораздо меньшую величину, чаще всего, как оказалось – 8 Гигабайт⁴. В моём случае всё было ещё хуже – MegaRAID вообще отдавал DOS-у ёмкость всего в 1 Гигабайт⁵! Забавно, что при этом fdisk от MS-DOS 6.22 показывал наличие неизвестного типа раздела на нём размером... в 8 Гигабайт. И доступ к MegaRAID осуществлялся через старый INT13, тогда как Adaptec 2100S «показывал» объём через extended INT13 (но те же 8 Гигабайт). Это означало, что, работая в DOS-ской программе, я не смогу склонировать свои существующие ~27 Гигабайт на новые ~140 Гигабайт – система до них просто не «дотянется». Мои прежние успешные опыты с NG4NW производились на одиночных SCSI-дисках, копировать RAID-массивы этой программой ранее мне не доводилось.

Для меня выбор сузился до одной программы – Storage Manager от фирмы Portlock. При изучении документации стало понятно – это именно то, что нужно. Программа умеет работать как из «голового» DOS-а, так и из Netware. В последнем случае, естественно, ей доступны все дисковые объёмы, которые «видит» Netware. Итак: Portlock StorageManager. Использованная версия – 3.27 от февраля 2004 года, хотя с NW4.11 вполне могла справиться и версия на несколько лет «постарше»⁶. Программа была скопирована на том SYS: в директорию по умолчанию SYS:STORMGR, а также на досовский раздел, в C:\STORMGR.

Несколько удивило отсутствие в StorageManager-е прямого копирования (клонирования) с одного диска (массива) на другой. В NG4NW такая функция наличествовала. Это означало, что копировать придётся через сохранение промежуточного образа сервера, с последующим его восстановлением на новый массив. Имея в виду вышесказанное про «хрупкость» конфигурации MegaRAID-а, наличие обязательной полной промежуточной копии образа сервера можно считать скорее благом, чем недостатком. Недостаток в этом случае только один – полное время выполнения всей операции переноса (фактически оно удваивается).

⁴ Понять это можно – зачем DOS-у больше?

⁵ Скорее всего это ограничение происходит от очень давно придуманного ограничения видимой для DOS ёмкости дисков, предоставляемой SCSI-контроллерами. Массив MegaRAID в этом смысле, видимо, просто изображает из себя один большой SCSI-диск.

⁶ Версия 3.27 декларирует поддержку всех типов файловых систем - и TFS и NSS для Netware всех актуальных версий системы, вплоть до 6.5.

Storage Manager предоставляет большой выбор для приёмника копируемого образа: это может быть лента, CD-R(W)/DVD-R(W) диски, ftp-сервер, специальная программа, запущенная на другой машине и по tcp/ip принимающая файл образа и др. Всегда желательно иметь как можно меньше потенциальных точек отказа, поэтому вариант с отдельной машиной-приёмником был отвергнут (исключается ещё одна машина, промежуточное сетевое оборудование и т.п.). В качестве рабочего был выбран вариант с DLT магнитофоном (DLT7000), подключаемым через отдельную плату SCSI-адаптера Adaptec АНА2940 на время проведения клонирования.

Копируем

Итак, суббота, можно надолго оккупировать предмет разборок, хозяева (бухгалтерия) звонить по поводу неработающего сервера не будут. Сервер останавливается, в него вставляется плата АНА2940, к ней подключается стриммер DLT7000⁷. Загружаю сервер. Здесь был вариант загрузиться без выполнения autoexec.ncf (server -na), но я запустился как обычно, а потом просто выгрузил лишние модули, опасаясь, что старт с ключом -na не загрузит чего-нибудь нужного. Дальше я запустил собственно Storage Manager: "load c:\stormgr\stormgr"⁸, добрался до главного меню программы и, переключившись на окно консоли, **руками размонтировал все тома сервера!** Иначе перед созданием образа SM сам размонтирует тома, а по окончании копирования попытается их смонтировать вновь, а нам этого совсем не надо. Дальше из меню выбрал операцию создания образа: Image -> Image server. В возникшем списке кандидатами на копирование по умолчанию отмечены все тома сервера и DOS-овский раздел. Это именно то, что мне требуется в данном случае. В следующем меню необходимо указать приёмник для создаваемого образа (Select Image Destination). Как уже отмечалось выше, в качестве такового я решил использовать магнитофон DLT7000, поэтому нужный пункт меню – «Write image to a tape drive». Я отказался от предлагаемой программной компрессии при записи на ленту, т.к. это (по идее) дополнительно нагрузит и без того не очень быстрый процессор старого сервера, а, главное – и без сжатия все содержимое дисков помещалось на одну ленту (40 Гигабайт). Дальше необходимо было указать конкретную модель из списка подключённых к серверу магнитофонов. У меня таковых был ровно один, поэтому особо выбирать было не из чего. Важный момент: какой способ записи использовать – с рекордами постоянной или переменной длины. DLT «любит» «Variable Length records», поэтому я его и выбрал (это, вообще-то, хорошо и толково расписано прямо в меню программы SM). Всё, жму «Enter» и созерцаю на экране начавшийся процесс копирования на ленту. Процесс достаточно долгий, на моём «железе» он шёл со средней скоростью ~4-5 Мбайт/сек. Общее время, требуемое на запись, легко вычисляется, его достаточно, чтобы покурить, выпить чашечку кофе или заняться другими, не менее полезными делами.

Когда контроллер нового RAID-массива впервые вставлялся в сервер, то соответствующий ему драйвер (в моём случае это был стандартный новеловский I2O) уже был скопирован в C:\NWSERVER и его загрузка было включена в STARTUP.NCF. Поэтому, дождавшись окончания успешного копирования всех «старых» томов сервера на ленту, я мог сразу же приступить к процедуре восстановления образа с ленты на новый массив. Хотя ничему бы не противоречила подгрузка I2O драйвера именно в этот момент, дело вкуса.

⁷ Заранее эту процедуру проделать было невозможно, т.к. DLT7000 штатно работает на другом сервере (сервер резервного копирования), который «отдыхает» только в выходные дни.

⁸ Для Netware 4.11 SM имеет два варианта работы с дисковыми накопителями – в «старом» стиле (DSK драйверы) или с использованием NWPA (HAM). В моём случае все драйверы были NWPA, поэтому, естественно, выбран был именно этот метод общения с дисками.

Запускаю процедуру восстановления (пункт Restore из главного меню программы Storage Manager). Как известно, в Netware, даже в размонтированном виде, не имеют права существовать тома с одинаковыми именами. Поскольку восстановление делалось без перезагрузки и без физического отключения старого массива, то было необходимо дать другие, не конфликтующие имена для вновь создаваемых томов на новом массиве. Конкретные имена новых томов роли не играют, я выбрал простую схему, просто добавив к реальному имени суффикс `_NEW` (`SYS_NEW`, `SOFT_NEW`, `DATA_NEW`). К небольшому сожалению, восстановление делается по одному тому за заход и новое имя можно дать восстанавливаемому тому только непосредственно перед началом процедуры копирования с ленты на диски. Поэтому операция restore, в отличие от создания образа, требует присутствия возле работающего сервера, т.е., курить придётся с перерывами **J**.

При восстановлении томов заодно были поправлены (увеличены) их размеры, основное увеличение пришлось на том `DATA_NEW` – с прежних ~15.5 Гигабайт он был увеличен до ~120 Гигабайт.

Закончив успешное восстановление и, по-прежнему **не перегружаясь**, я переименовал все старые тома в имена с добавленным суффиксом `_OLD`: `SYS_OLD`, `SOFT_OLD`, `DATA_OLD`. У новых (восстановленных) томов, соответственно, в именах был убран суффикс `_NEW` и они получили окончательные рабочие имена: `SYS`, `SOFT` и `DATA`. Переименование делалось средствами Storage Manager-а, из меню работы с томами.

Осталось совсем немного – отключить и «разобрать» старый массив, и включить загрузку с нового массива. Делаю серверу `down->exit`, перегружаюсь и, в нужный момент, жму заветные `Ctrl-A`, чтобы в очередной раз попасть в `setup BIOS`-а контроллера Adaptec 2100S. Я помнил, что при первой попытке входа в этом месте возникла большая задержка, около четырёх минут, поэтому, не увидев мгновенной реакции на `Ctrl-A`, я приготовился ждать примерно это же время. Но когда картинка настроек на экране не возникла через пять, восемь, десять (!) минут – я стал сомневаться – всё ли у меня в порядке с железом? Индикация на 2100S опять демонстрировала некоторую активность, но лампочки дисков практически безмолвствовали. Что на этот-то раз не так? Документированная задержка связки нашей MB440LX и Adaptec 2100S – да, это у меня в памяти отложилось, но не десять же минут! Как оказалось – шестнадцать! Примерно через четверть часа, когда рука уже так и тянулась нажать кнопку сброса, на экране наконец-то возникла нужная картинка. Однако! А ведь менее терпеливый мог запаниковать и устроить разборки на пустом месте, это ж форменный бардак – выдерживать измученного ожиданием админа по четверти часа перед заходом в `setup J`.

Дальше всё было просто – в `setup`-е была включена поддержка загрузки с 2100S, после чего сервер перезагрузился в очередной раз, с нажатием кнопки F8 на старте, чтобы пройти процедуру загрузки по шагам (и не грузить `server.exe`). Оказавшись в `DOS`-е, решил проверить – а с какого массива сервер всё-таки загрузился? Как и положено, `DOS` «видел» два (вместо одного) диска – `C:` и `D:`, естественно, абсолютно одинаковых (клон!). Да, конечно, я помнил, что когда «пустой» 2100S был впервые установлен в сервер, загрузка пыталась идти именно с него (из-за чего её тогда пришлось временно отключать), но сейчас хотелось явно и наглядно убедиться – система грузится с 2100S, а не с MegaRAID. Решение было простым и весёлым – средствами `VC`, который я всегда держу на досовских разделах всех своих серверов, был запущен поиск какой-то бессмысленной строки по всем

файлам по всему С:. Индикаторы при этом активно заморгали именно у дисков, расположенных на «нижнем этаже», т.е., на новом массиве **J**.⁹

Т.к. новый массив уже полностью готов к загрузке, я закомментировал загрузку драйвера старого MegaRAID-а в STARTUP.NCF (строка LOAD MEGA4_XX.HAM SLOT=10003 NS=1), чтобы, запустившись, сервер не стал автоматически монтировать все обнаруженные тома, включая старые.

Теперь самое время вынуть плату АНА2940, через которую был подключён магнитофон DLT7000. Просто выключая питание сервера (был загружен DOS), вытаскиваю плату, запускаю сервер в обычном режиме, слежу за загрузкой – всё в порядке, все службы запустились, дерево ожило, пошла синхронизация.

Последнее действие – надо наконец-то «разобрать» старый массив и снять четыре девятигигабайтных диска с «верхнего этажа»¹⁰. Ломать – не строить, это мы сейчас быстро состроим **J**. Подгружаю руками (надеюсь – в последний раз!) драйвер MegaRAID-а, запускаю его утилиту управления и контроля: LOAD MEGAMGR, начинаю искать в её меню метод убийства массива. Ха-ха три раза – не могу найти! До сих пор все разборки с этим массивом были ориентированы исключительно на создание и ремонт массива, а вот чтобы разрушить – это у меня с ним впервые. Инициализация? Нет, это долго, да и не то, что нужно – просто будет создан новый пустой массив. А мне его надо аннулировать совсем. Смешно, но поиск нужного пункта занял минут пять, не меньше. Верным оказался пункт “Clear Configuration” из меню “Configure”. Век живи... В общем, это уже точно всё – очистил конфигурацию MegaRAID-а, по одному, аккуратно, извлёк все четыре старых диска, выгрузил драйвер MEGA4_XX.HAM. Сервер работает с новым массивом, никто ничего не заметил, если не считать существенным пятикратное увеличение дисковой ёмкости и простой сервера в течение нескольких часов. Ура!

Небольшое послесловие

У меня в хозяйстве есть ещё один, практически точно такой же сервер (“Madrona”), которому в ближайшее время предстоит такая же процедура переезда со старого дискового массива на новый. Контроллер нового массива и новые диски абсолютно такие же, как в описанном выше случае, но вот старое железо имеет некоторые отличия, из-за которых повторить проделанную процедуру один-к-одному, не задумываясь – не удастся. Причина простая и грустная одновременно – в этом сервере (он на полгода «постарше») диски в MegaRAID массиве использованы другие – не четыре по девять гигабайт, а шесть штук по четыре. Поскольку каждая из «корзин» сервера вмещает не более пяти дисков, это означает, что фокуса с одновременным сосуществованием в сервере старого и нового RAID-контроллеров, когда каждый из них обслуживал собственный «этаж» со своими дисками, проделать не удастся – невозможно «переселить» все старые диски в одну «корзину». Поэтому на этом сервере клонирование будет разбито на несколько этапов – вначале содержимое старого массива будет записано на ленту, потом старый массив будет разобран (т.е., в этот момент сервер останется «жить» только на ленточке. Страшно! **J**), затем диски нового массива (они уже собраны в массив заранее) будут установлены в «корзину» и... вот дальше начинается самое интересное.

⁹ Приятная особенность полного «клонирования» DOS-раздела – на новом массиве он также автоматически получил статус «Active», унаследованный со «старого» DOS-а, и никаких дополнительных манипуляций с FDISK-ом и т.п. не потребовалось.

¹⁰ Диски можно снимать «на ходу», это hot-swap модели, поэтому дополнительной остановки сервера для этой процедуры не требуется.

Как уже упоминалось выше, в системе DOS Adaptec RAID 2100S отдаёт размер массива не более 8 Гигабайт. Это означает, что восстановить с ленты все ~20 гигабайт на ~140 гигабайтный массив невозможно. Portlock рекомендует в этом случае поступить следующим образом – если раздел DOS + том SYS не превосходят вместе 8 гигабайт, то надо восстановить вначале досовский раздел и том SYS из DOS-версии Storage Manager-a, затем загрузить с них минимальный сервер и продолжить восстановление остальных томов уже из Netware. Но я думаю поступить немного иначе: подключить к серверу на время восстановления какой-нибудь IDE-диск с установленной на нём Netware 4.11 (и с DOS-ом, естественно), загрузиться с него, запустить сервер NW4.11 в минимальной конфигурации **без** монтирования томов и восстановить с ленты сохранённые тома и DOS-раздел одним заходом. Мне кажется, что такой вариант восстановления всех разделов с использованием run-time Netware будет проще. Кстати, мелочь, а приятно – не потребуется процедуры переименования томов туда-обратно, ведь когда будет идти восстановление, старого массива в сервере уже не будет, конфликты по именам не возникнут.

*Сергей Дубров,
20 апреля 2004 года*